A Findings Model for an Ambulatory Pediatric Record: Essential Data, Relational Modeling, and Vocabulary Considerations

Richard N. Shiffman, M.D., M.C.I.S.
Center for Medical Informatics and the Department of Pediatrics
Yale School of Medicine, New Haven, Connecticut

Effective, computer-based representation of clinical observations requires balancing the advantages of structured, coded descriptions against those of free text narrative. An essential data set of relevant signs and symptoms was defined by a multidisciplinary group based on management goals published in a national guideline to meet the needs of clinicians in the Spina Bifida Clinic at Yale-New Haven Hospital. These core data elements are stored in a structured format. Additional material is stored as free text.

A relational schema was devised that permits storage of both coded findings and narrative. Symptoms and signs are represented as subtypes of a supertype patient finding entity; they inherit common attributes and specialize others.

The IVORY vocabulary was supplemented and modified to provide terms that describe relevant clinical observations. For this application, fields were added that enable predictive data entry of findings based on patient age and gender.

INTRODUCTION

A major challenge that faces the designers of a computer-based medical record is to effectively represent the clinical data that are captured during a health care encounter. Free-text narrative, the conventional representation in the paper-based record, is flexible, expressive and familiar to clinicians [1]. With current technology, however, computers are able to make only limited use of information stored in a free-text format [2]. On the other hand, coded data—precisely defined and based on controlled vocabularies—can be used to organize the medical record for efficient search, retrieval, and display, to aggregate data for research and administrative functions, and to trigger decision support. However, coded data may be insufficiently expressive and structured data entry may be unwieldy.

This paper describes an approach that was applied to balance these representations in a computer-based ambulatory record. A core data set was prospectively identified that is maintained in coded format to meet the needs of clinicians, followed by those of researchers and administrators. Additional information is maintained in free-text format. We describe the process for choosing the core data sets, the development of a relational architecture for storage of these data, and the selection of a controlled vocabulary for representation of the core data.

Clinical Environment

The Department of Pediatrics at the Yale School of Medicine has established a center for the care of Children with Special Health Care Needs (CSHCN). It provides diagnostic and therapeutic services to pediatric patients who have a broad range of disabling conditions, including spina bifida, cerebral palsy, congenital and acquired heart disease, cleft lip and/or palate, rheumatoid arthritis, and cystic fibrosis. A replicable and extensible information architecture was needed to manage the clinical, research, and administrative data generated in caring for children with these chronic diseases. A prototype clinical information system has been developed using Microsoft Access Version 2.0 (Microsoft Corp., Redmond, Wash.) with a plan to migrate to a client-server architecture.

Initially, the system is being implemented in the Spina Bifida Clinic at Yale-New Haven Hospital. Comprehensive care of children who have spina bifida—a congenital disorder that results from abnormal intrauterine development of the central nervous system and spine—requires coordination of multiple disciplines, including neurosurgery, orthopedics, urology, pediatrics, nursing, social services, physical and occupational therapy, orthotics, and genetic counseling. The clinical information requirements in this clinic are therefore quite extensive and should be representative of the needs of other ambulatory care environments.

CORE DATA SET

Several initiatives in the United States and Europe have focused on the specification of basic information to support clinical services. For the ambulatory setting, the Uniform Ambulatory Care Data Set defines a common core of data items and provides standard definitions [3]. Likewise, the Advanced Information in Medicine Program of the European Economic Community has defined minimum data sets to characterize clinical information requirements and constraints and to enhance uniformity [4]. Both are predicated on an assumption that there is a core of data common to the needs of multiple users that should have a priority for standardization. However, neither of these multi-purpose data sets is sufficiently detailed to adequately describe the findings observed during a clinical encounter.

The American Society of Testing and Materials Standard ASTM E1384-91 identifies the essential content and logical structure of an electronic primary health record [5]. The data elements included are extremely broad-based and appear comprehensive, but components that define clinical observations during encounters are minimal.

Moidu et al. suggest that an *essential data set* should include "just as many data elements as required to provide the essential foundation for decision-making...to identify the need for care, [and] to monitor and evaluate the impact of the care provided" [6]. This definition was applied to prospectively define data relevant to clinical management and outcomes assessment for the CSHCN program.

Use of Guidelines to Define Core Data

Well-crafted guidelines represent a valuable knowledge resource. They provide up-to-date information that has been sanctioned by the sponsoring organization. Such information can be used to break the knowledge acquisition bottleneck for the development of knowledge-based systems and can serve as a basis for decision support [7]. In this work, knowledge from a published practice guideline was reused to define an essential data set for an ambulatory patient record [8].

The Spina Bifida Association of America has published guidelines for the management of patients with spina bifida, which are intended to maintain optimal health status, prevent secondary disabilities, maximize the potential to participate in society, and foster independence [9]. The guidelines provide age- and discipline-specific recommendations to health care providers.

For example, the Neurosurgical guidelines from infancy through adolescence focus on 3 major goals:
• maintenance of normal intracranial pressure
• recognition of presence of the Chiari malformation
• recognition of evidence of cord-tethering symptoms
Clinical management pertinent to these goals requires regular consideration and documentation of a significant number of historical findings and examination observations.

Comparatively little attention has been paid to data management for the most fundamental observations made by clinicians—information elicited in the clinical history and physical examination. It is estimated that these findings lead to a diagnosis in 90% of cases, without requiring the expense and discomfort associated with supplemental laboratory tests and imaging procedures [10]. Findings data are also essential to characterize clinical outcomes, an area of increasing concern in the current managed care environment.

For the neurosurgical component of the Spina Bifida Clinic database, representatives from neurosurgery, nursing, and pediatrics identified 17 symptoms and 22 signs that are directly relevant to the above-mentioned clinic goals. It was agreed that these data elements would be coded in a structured manner. Additional findings data that are collected during clinical encounters can be captured as free text and are available in clinical reports, but they cannot be aggregated or used for record organization.

RELATIONAL DESIGN

The Canon group has focused collaboratively on the representation of medical concepts related to radiography [11] and their publications have concentrated on that domain, e.g. [12, 13]. Other efforts to model clinical observations have dealt with the realms of laboratory medicine [14], and endoscopy [15]. However, these publications are only indirectly applicable to pragmatic issues of database design for symptom and sign data.

Recently, Dolin has presented a high-level conceptual model of symptoms that uses a nested, polyhierarchical representation [16]. His use of a single table to store symptoms, modifiers, and relationships is simpler than a conventional, multi-table design but manipulation of the data is more complex.

Structured Coding of Findings Data

The approach described by Fleming and von Halle was applied to the logical design of a findings model for the Spina Bifida Clinic database [17]. This methodology systematically analyzes and graphically models information requirements and translates the model into a stable relational implementation. The procedure is data-driven, i.e., it is based on an understanding of how the information is used without consideration of specific processing patterns. Clear, graphical diagrams are produced that facilitate communication among designers, developers, and users. The end-result is a consistent, sharable, and flexible database.

In this report, attention is focused on modeling of symptom and sign data. A skeletal user view is created to define and model the major relevant data objects—initially without supportive detail. The entities of interest in this model include PATIENTs, PROVIDERs, ENCOUNTERs, and the FINDINGs themselves. SYMPTOMs and SIGNs are subtypes of the supertype entity FINDING (as are LAB RESULTs and IMAGING RESULTs). Symptoms represent phenomena experienced subjectively by the patient and reported by the patient or a surrogate historian. Signs are objectively observed by clinicians. Each of these subtype instances represents the same "object" in the real world as the supertype;

Patient

PatientID
PatientName
DOB
PatientAddress
Gender
...

Provider

ProviderID
ProviderName
Provider Type
Address
Phone
...

Encounter

EncounterID
PatientID
ProviderID
EncounterDate/Time
Type_of_Service
...

Patient_Finding

Patient_FindingID
EncounterID
FindingID
Comment
Context
Certainty
Absent
Comparison operator
Comparison finding
EnteredBy
TimeStamp

To Findings Vocabulary

Symptom

Patient_FindingID
Onset
Past History
Duration
Freq
Trend
Intensity

Location

Patient_FindingID
Location
Laterality

Quality

Patient_FindingID
Quality

Radiation

Patient_FindingID
Radiation

Elicited_By

Patient_FindingID
Brought_on_by

Relieved_by

Patient_FindingID
Relieved_by

Sign

Patient_FindingID
Measured Value
Finding Unit
Intensity

Laboratory

Patient_FindingID
Finding Value
Finding Unit
Facility
ReportedBy
...

Imaging

Patient_FindingID
Modality
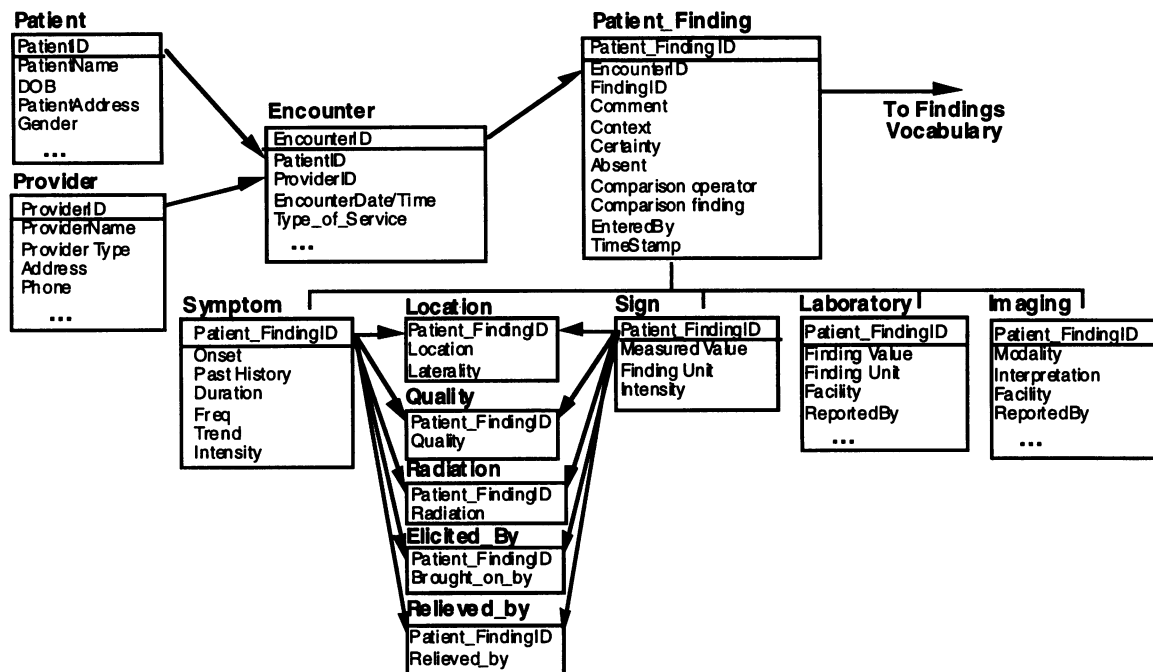Interpretation
Facility
ReportedBy
...

Figure 1. Logical model of findings for the Spina Bifida Clinic database.

each subtype instance has all the properties of the supertype plus some additional properties; and, for each instance of the subtype, there exists exactly one instance of FINDING, although the converse need not be true.

Relationships link these major entities. PATIENTs and PROVIDERs are linked one-to-many to ENCOUNTERs. ENCOUNTERs and FINDINGs are related many-to-many. Therefore, an associative entity (PATIENT_FINDING) is defined that relates one-to-many with both ENCOUNTER and a FINDING list. SYMPTOMs, SIGNs, LAB RESULTs, and IMAGING RESULTs represent mutually exclusive categories of PATIENT_FINDING, each associated 1:1 with the supertype.

Each instance of these entities is defined by a key, a unique numeric identifier. The primary key of each subtype entity is the same as that of the supertype FINDING entity.

The next step in the creation of the logical model is to add non-key attributes to each entity (Figure 1). Fleming and von Halle recommend that non-key attributes be placed as high as possible in the logical model; following normalization principles, each must be fully determined by the entire primary key.

Non-key attributes for PATIENT and PROVIDER entities are implementation specific and beyond the scope of this paper. ENCOUNTERs are typically described by a type of service and a site of service, as well as provider ID, patient ID, and Date-time (an alternate, composite key).

Non-key attributes for the FINDING entity are common to all occurrences of the supertype. General, non-key attributes that are pertinent to all findings include:

• Finding ID: a pointer to the clinical vocabulary term list (see below)
• Comment: this attribute permits free-text description to capture the richness of the clinical encounter. This text field can be used to modify specific findings, e.g., Rector's example of angina evoked by walking past the frozen food counter, or it can store narrative not associated with any coded data item.
• Context: *abdominal pain* may occur both as a SYMPTOM reported by the patient and as a SIGN elicited by the examiner in response to deep or superficial palpation. Likewise, patients may report *wheezing* or clinicians may detect it on auscultation of the chest. The context field identifies the specific source of the finding.
• Absent: this Boolean attribute allows expression of negation
• Certainty: this attribute expresses various levels of certainty, e.g., probable, possible, $P=.75$
• Comparison with: this attribute stores the identifier of another finding with which the current finding is compared
• Comparison operator: the relationship of the current finding to the comparison finding, e.g., smaller than, milder than, >=, etc.
• EnteredBy, TimeStamp: The identifier of the user and the time the finding is recorded.

413

The SYMPTOM subentity inherits all of the above attributes from the FINDING entity and adds specialized attributes. A primary difference between symptoms and signs is the fact that symptoms are described by temporal modifiers. These attributes include:

- Onset: e.g., abrupt, indolent, one day, 2 weeks
- Past history: references previous episodes of a SYMPTOM, e.g., 1 year ago
- Duration: the length of time a SYMPTOM continues
- Frequency: e.g., rare, constant, seasonal
- Trend: unchanging, improving, worsening
  Both SYMPTOMs and SIGNs are modified by:
- Intensity, e.g., moderate, disabling, copious, Grade II/VI,
  THE SIGN entity adds specializations for:
- Measured values
- Units

e.g., 120 (Measured value) millimeters of mercury (Units).

Some SYMPTOM and SIGN attributes are multi-valued, i.e., a given FINDING may take on several modifier values simultaneously. These attributes are classified as child entities:

- Location: a modifier may be necessary when the finding does not connote location, e.g., a *rash* may be present on the right arm and left leg.
- Laterality; a topographic modifier, e.g., left, right bilateral
- Quality: e.g., red, papular, matted, coarse
- Radiation: e.g., chest pain can radiate to the left arm and the left jaw
- Elicited by: e.g., meals, 1 flight of steps, deep palpation, Ortolani's maneuver, caloric stimulation
- Relieved by: e.g., sitting up, OTC analgesics, bowel movement.

The logical model depicted in Figure 1 can be readily transformed into a relational implementation. Tables are created to represent each entity. Columns in these tables represent each attribute. Indexing, redefinition of columns and tables, denormalization, and other "tuning" procedures may then be performed to meet performance requirements.

VOCABULARY

Encoding data related to symptoms and signs has been challenging because the concepts are often complex and a comprehensive controlled vocabulary has been lacking. We modified the IVORY vocabulary for use in this application [18]. This vocabulary, originally derived from the Wisconsin Ambulatory Research Project, was modified by Campbell and Musen for use in their IVORY progress note tool [19].

The IVORY vocabulary includes unique identifiers for 966 findings terms and a large number of modifier terms grouped by categories such as

location, frequency, severity, and aggravating factors. Findings terms are linked by associative tables to relevant modifiers. Many of the modifiers match attributes in the relational model described here. The expressiveness of this vocabulary makes it a reasonable choice for basic terminology.

For this application, predictive data entry "forms" were designed to display only the symptoms and signs relevant to a given patient's age group and gender. This necessitated adding attributes to the vocabulary terms (Figure 2). With human growth and development, the symptoms and signs that indicate the outcomes of interest change. For instance, symptoms of increased intracranial pressure in infancy include irritability, lethargy, and vomiting. Signs include a full anterior fontanel, accelerating head circumference growth rate and sunset sign. On the other hand, older children may complain of headaches and suffer from memory disturbance, blurred vision, and decreased school performance. Additionally, the fontanel closes, head circumference does not change in response to acute pressure changes, and papilledema is more easily ascertained. Similarly, gynecologic terms are pertinent only when the patient is female.
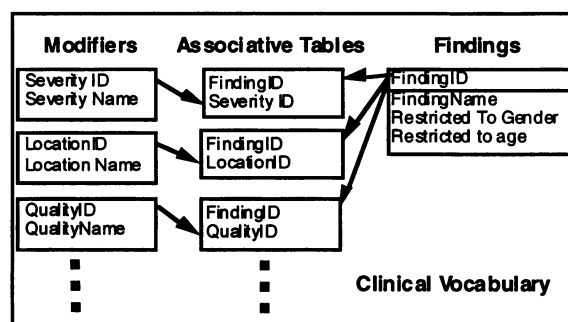

Figure 2. Relational structure of the modified controlled vocabulary.

Such restrictions are easily added as fields to the IVORY findings vocabulary. The fact that the finding identifier carries no contextual information (as it does in hierarchical classification systems such as ICD-9-CM and SNOMED) facilitates addition of new terms.

As noted by Lindberg and Humphreys, "There are no existing controlled vocabularies that can be recommended without reservation for even selected elements of the automated patient record" [21] The IVORY vocabulary lacks many of the terms necessary for the Spina Bifida Clinic application. Of the 43 terms required to encode core data for the neurosurgical component of the database, the IVORY vocabulary lacks terms for 22 and has marginal semantic matches for another 2.

Moorman et al. caution that the use of a limited set of modifiers prohibits domain completeness and results in a limited descriptive level of detail [20]. A facility is included in the database that allows the

option to add terms for modifiers. While initially coded as "Other", such terms can be reviewed and added to the vocabulary.

The recent release of version 3.1 of SNOMED International (College of American Pathologists, Northfield, Illinois) promises access to a rich, standardized nomenclature of concepts for symptoms and signs. It contains more than 800 newly added signs and symptoms. Of the neurosurgical concepts required for the Spina Bifida Clinic database, SNOMED lacks terms for only two (decreased school performance and sunset sign). SNOMED codes can be added as attributes of the IVORY vocabulary.

FUTURE DIRECTIONS

Each of the components described above will be extended as this clinical database is fully implemented. Additional disciplines will define essential data sets and the adequacy of this approach for selection of core data will be evaluated. The relational model of findings will be extended to allow findings to support meta-level problems, and assessment and plan entities will be added. Necessary terminology will be added to the vocabulary and linkages to SNOMED will be enhanced.

References
1. Dick RS, Steen EB, ed. The computer-based patient record: an essential technology for health care. Washington, D.C.: National Academy Press, 1991
2. Rector AL, Nowlan WA, Kay S. Foundations for an electronic medical record. Meth Inform Med 1991;30:179-86.
3. Report of the National Committee on Vital and Health Statistics and the Interagency Task Force on the Uniform Ambulatory Care Data Set. U.S. Dept. of Health and Human Services, 1989.
4. Parkin D, Hutchinson A, Fisher PJ. Minimum data sets for ambulatory care. In: O'Moore R, et al., ed. MedInfo Europe. Glasgow, Scotland: Springer Verlag, 1990: 88-93.
5. ASTM. Standard guide for description for content and structure of an automated primary record of care E1384-91. 1991.
6. Moidu K, Singh AK, Boström K, et al. Towards an essential data set: applicability in the domain of maternal health services. Meth Inform Med 1992;31:182-92.
7. Shiffman RN. Towards effective implementation of a pediatric asthma guideline: integration of decision support and clinical workflow support. In: Ozbolt J, ed. Proc 18th SCAMC. Washington: 1994: 797-801.
8. Musen MA. Dimensions of knowledge sharing and reuse. Comput Biomed Research 1992;25:433-67.
9. Rauen K. Guidelines for Spina Bifida Health Care Services Throughout Life.Washington, DC: Spina Bifida Assocaition of America, 1990.
10. Lipkin M. The care of patients: perspectives and practice. New Haven, Conn.: Yale University Press, 1987:82.
11. Cimino JJ. Controlled medical vocabulary construction: methods from the Canon Group. J Am Med Informatics Assoc 1994;1:296-7.
12. Friedman C, Cimino JJ, Johnson SB. A schema for representing medical language applied to clinical radiology. J Am Med Informatics Assoc 1994;1:233-48.
13. Bell DS, Pattison-Gordon E, Greenes RA. Experiments in concept modeling for radiographic image reports. J Am Med Informatics Assoc 1994;1:249-62.
14. Stahlhut RW, McCallie DP, Waterman DM, et al. A relational model for clinical objective results. Proc 14th SCAMC 1990:354-58.
15. Gouveia-Oliveira A, Salgado NC, Azevedo AP, et al. A unified approach to the design of clinical reporting systems. Meth Information Medicine 1994;33:479-87.
16. Dolin RH. Modeling the relational complexities of symptoms.Meth Inform Med 1994;33:448-53.
17. Fleming CC, von Halle B. Handbook of relational database design. Reading, Mass.: Addison-Wesley, 1989
18. Campbell KE, Musen MA. Creation of a systematic domain for medical care: the need for a comprehensive patient-description vocabulary. In: Lun KC, et al., ed. MEDINFO 92. Elsevier Science Publishers B.V. (North-Holland), 1992: 1437-42.
19. Campbell KE, Wieckert K, Fagan LM, et al., A computer-based tool for generation of progress notes. In: Safran CS, ed. Proc 17th SCAMC Washington: 1993: 284-8.
20. Moorman PW, van Ginneken AM, van der Lei J, et al. A model for structured data entry based on explicit descriptional knowledge. Meth Inform Med 1994;33:454-63.
21. Lindberg DAB, Humphreys BL. The Unified Medical Language System (UMLS) and computer-based patient records. In: Ball MJ, Collen MF, ed. Aspects of the computer-based patient record. New York: Springer-Verlag, 1992: 172.